IP Based Distributed Virtual SAN

By
Sheng (Ted) Tai Tsao

Field of the Invention

The present invention relates generally to IP based distributed virtual SAN, its automatic configuration, its volumes allocation and accessing.

Background Information

a) Basic Terminology:

SAN:

SAN is a storage system comprises multiple storage media such as disk drives and provides computer host client with block data through network media, which is either cable (Fibre-optical or regular cable) or wireless connected by using protocol such as TCP/IP/UDP, Fibre-Channel, or other protocols.

DNS:

Network domain name server, which help network station to find their target network address.

SNMP:

SNMP is an abbreviation for "Simple Network Management Protocol", which is a standard Internet protocol. The SNMP trap is a UDP packet sent by SNMP daemon on a SNMP agent system to SNMP network management station through network link.

b) Why Do We Need Virtual SAN:

Today's corporate IT professional faces many challenges to handle the ever increase information and data. This often requires many organizations to expand their storage capacity, manage storage systems and to keep the normal business run. Currently, The IP based NAS (network attached storage) effectively provides storage and services for end user's file system needs. On the other hand, at the enterprise level, the majority storage systems are still server directly attached and being accessed as raw block data devices through either traditional SCSI or Fibre Channel technology.

The server direct attached storage system has many drawbacks, which are described as follow:

a) Currently, the most advance storage management system only capable to handle 4TB of data, which is far from good enough for enterprise storage management requirement.
b) The most of server directly attached storage has problems to expand its capacity. In some case, it is quite often to require purchasing a new server in

order to expand the storage. In other cases, it also requires to shutdown the server and to stop the normal operation in order to expand the storage capacity.

c) The storage being attached can only be accessed by the attached server and can not be shared by other even a server's storage has spare capacity left while other server are in shortage of the storage capacity within a department or cross department in a organization.
d) Each attached storage system has to be managed separately and this is a nightmare for IT professional.
e) With the attached storage system, the backup/restore has to go through the data network, this will tax the data network performance.
f) The SCSI only allow 12 meter distance for data accessing with 15 storage devices while Fibre Channel is also limited to 10 kilometers long. This effectively prevents them from being the best choice for disaster recovery of the storage system.
g) The Fibre Channel based storage system cannot handle well for the interoperability. Also, Fibre Channel based storage system is expensive to build and to maintain.

Brief Description of Invention

With the rapid development of network technology such as wide adoption of Gig Bits (1GB and 10-GB) Ethernet technology, the problems mentioned above can be solved by built IP based Virtual SAN. IP based virtual SAN is a method to group multiple SAN units together through IP network technology to form a huge capacity storage system, which provides block data to multiple computer host clients. The benefits of building such IP based SAN are:

Scalability:

A hundreds or thousands terabytes raw block data pool can be built initially at storage system automatic configuration time. Afterwards, its capacity can be dynamically expanded without interfering the normal server accessing and storage operation by adding one or more IP SAN boxes based on demand. This could meet variety customer's needs from small to large.

Data Sharing:

Unlike server attached storage, which cannot be shared even there is larger percent of storage space unused. Within IP based virtual SAN, if a SAN box being configured and exported with multiple logical devices (volumes), each volume in SAN box can be accessed by a single server and different volumes can be accessed by different multiple servers concurrently. This allows departments in a corporate to fully sharing the storage.

Centralized Data Management:

Unlike the server attached storage, in which each storage has to be managed separately, with the IP based SAN all storages can be managed through a single centralized UI window on the Web for logical devices/volumes management such

as disk/raid configuration, partitioning and re-partitioning, security management, server storage allocation or de-allocation by client hosts and accessing management, fault handling management, data replication and backup/restore management, and all others.

Fault Handling:

Unlike the Fibre Channel, which is limited to 10 kilometers range, the IP based virtual SAN within corporate intranet can span cross the states, countries, and even continents. Therefore, a disaster recovery plan for IP based virtual SAN can cover far beyond 10 kilometers range. This provides a much safe for fault and disaster recovery. With a careful planed fault handling hierarchy, either a server goes down or a storage unit goes down can be well recovered.

Cost Saving:

Compare with Fibre Channel based storage technology, the IP based distributed virtual SAN is relatively inexpensive to build and easy to maintain, this will effectively reduce the cost for those customers, who use this type of storage system.

The present invention focuses on how to use multiple SAN unit to form a distributed virtual SAN, how to automatically configure and build a huge capacity distributed virtual SAN, and how to allocate and access storage volumes of the distributed virtual SAN. This invention will become understood with reference to the following description, claims, and accompanying figures.

Brief Description of Drawings

Fig. 1: Shows an example of simplified block diagram of distributed virtual SAN infrastructure, which includes:

a) Client host (**1**): It will utilize the block data provided by this virtual SAN for their needs such as to build file system on it or build a raw device based database on it.

b) Network infrastructure (**2**): It includes Switches/Routers/gateways, which are either cable or wireless connected with different type of connecting media to form LAN/WAN. These network connection and infrastructure provide data path between client host, Distribution control and management station, and IP SAN Units. In addition, these network may either be a private storage network island or a corporate storage network backbone, where the virtual SAN could be build up and at meantime allow client hosts to access each individual SAN unit. It could be LAN (local area network) or WAN (wide area network). The network infrastructure also includes software infrastructure such as DNS, which allow the distributed virtual SAN operated in a cross network domain environment.

c) Distribution control and management station (**3**): It controls and manages the entire virtual SAN such as to automatic configure virtual

SAN, to redistribute client hosts block data request to an individual SAN etc.
It contains distribution control management software modules and maintains a list of SAN unit information. Each entry in the list contains information such as unit name, unit IP address, unit storage information, unit status such as online or down, and more.

d) IP SAN unit (**4**): It is the actual SAN and provides block data to client hosts. It contains SAN service software modules, which provide services either to distribution control and management station or client hosts.
e) Fibre Channel to IP Gateway (**5**): It translates between Fibre Channel based protocol and IP based protocol so that Fibre Channel based SAN unit will appears as if IP based SAN unit to the rest of the world (Fig. 1).
f) Fibre Channel SAN Unit (**6**): similar to IP SAN unit except it uses Fibre Channel protocol to communicate with parties. It is the responsibility of Fibre Channel to IP protocol Gateway (**5**) to convert Fibre Channel protocol to IP protocol and vise versa.

Fig. 2: This figure is a portion of Fig. 1. It represents the actual virtual SAN. It is provided for the convenience of discussion of automatic configuring and building the IP based virtual SAN since during this part of process there is no client host involved. The DNS included in the switches/routers network infrastructure to indicate the importance of the DNS in the crossing domain environment for network communication. The actual DNS (domain name server) may be placed in distribution control management station or somewhere else in a station within the network infrastructure (2 of FIG 2 and 2 of Fig. 1).

Fig. 3: This diagram shows a protocol of virtual SAN automatic configuration and building as well as shutdown.

Fig. 4: This Diagram shows the protocol message format, which used by "Virtual SAN Automatic Configuration Protocol"

Fig. 5: This Fig. Shows the storage in an IP SAN unit, which may be further divided into multiple volumes and each volume may be further divided into multiple partitions.

Detailed Description of the Invention

1: Distributed Virtual SAN:

Fig. 2 Shows a simplified diagram of a distributed virtual SAN according to this present invention. The distributed virtual SAN comprises one or more SAN units (4 of Fig. 2) connected to a distribution control management station (3 of

Fig. 2) via one or more switches or routers and other network infrastructure (2 of Fig. 2) described in previous section. Fig. 1 shows that one or more client hosts (1 of Fig. 1) can either connecting to distribution control management station (3 of Fig. 1) for requests of block data service or connecting to any IP SAN unit (4 of Fig. 1) for actual block data accessing after granted by distribution control management station (3 of Fig. 1). The IP SAN unit (4 of Fig. 1) may hold multiple storage volumes in the form of block data for client hosts (1 of Fig.1) accessing. This will allow multiple client hosts (1 of Fig. 1) to share an IP SAN unit (4 of Fig. 1) by granting and assigning each client host (1 of Fig. 1) to exclusively access particular volumes on that IP SAN unit (4 of Fig. 1). The distribution control management station (3 of Fig. 1) maintains each IP SAN unit's (4 of Fig. 1) information in a list.

2: Automatic Configuration:

The automatic configuration of distributed virtual SAN (Fig. 2) occurred during each IP SAN unit (4 of Fig. 2) being brought to online and the virtual SAN is being built up. The auto configuration also occurs at each IP SAN unit shutdown time during which the distributed control management station (3 of Fig. 2) update the virtual SAN configuration information. The Fig. 3 shows the Virtual SAN Automatic Configuration Protocol, which leads to the success of the construction of the distributed virtual SAN (Fig. 2) according to this invention. The network infrastructure represented by switches/routers (2 of Fig. 2) and others does not display in Fig. 3. However, the DNS included in network infrastructure (2 of Fig. 2) plays significant hidden role in this protocol (Fig. 3) due to DNS can help communication sender, which is either IP SAN Unit (4 of Fig. 3) or distribution control management station (3 of Fig. 3), find the address of the destination during sending messages in a crossing domain environment. This helps this distributed virtual SAN overcome the geometric region limitation. In addition, Fibre Channel SAN unit's (6 of Fig. 2) will appears as an IP based SAN unit to this distributed virtual SAN once it connects to a Fibre Channel to IP gateway (5 of Fig.2). Therefore, it will be treated the same as IP SAN unit in all of following discussion without additional comments.
The following steps have described the sequence of virtual SAN automatic configuration, which confirm to the Fig. 3. In addition, the role of DNS will not be mentioned in the following sequence due to it has been clarified already.

a) When any of IP SAN unit (4 of Fig. 2) such as unit (n) brought up to online, its SAN service modules sent out a "SAN unit (n) startup" packet to distribution control management station (3 of Fig. 2). This message could be a simple user defined UDP packet (Fig. 4) with message type of system up. This message also could be a SNMP trap of cold start packet if the IP SAN unit (n) was power off before, or could be a SNMP trap of link up packet if there was previous network link down on that IP SAN unit (n) (4 of Fig. 2).
b) When distribution control management modules of distribution control management station (3 of Fig. 2) receives IP SAN unit (n)'s message, it puts

the IP SAN unit (n)'s information such as unit name and unit IP address into the IP SAN unit information list of distribution control management station (3 of Fig. 2).

c) After putting information into the IP SAN unit information list, the distribution control management modules on distribution control management station (3 of Fig. 2) sends out a "need SAN unit (n)'s storage info" packet to IP SAN unit (n) (4 of Fig. 2).
d) When SAN service modules on IP SAN unit (n) (4 of Fig. 2) received packet of "need SAN unit (n)'s storage info", it gets storage information on IP SAN unit (n) (4 of Fig. 2), which includes the number of storage volumes, each volume's start address (logical block address, LBA), length, and the end address (logical block address, LBA). The SAN service modules then send a packet of "unit (n) storage info", which includes all information it obtained to distribution control management station (3 of Fig. 2).
e) After receiving "unit (n) storage info" packet from IP SAN unit (n) (4 of Fig. 2), the distribution control management modules on distribution control management station (3 of Fig. 2) updates its IP SAN unit information list with corresponding storage information of IP SAN unit (n) from packet.

After all IP SAN unit (4 of Fig. 2) are brought into online, the automatic configuration of the virtual SAN (Fig. 2) has finished. Further, the distribution control management station (3 of Fig. 2) has controlled entire virtual SAN since it owns storage volumes information and network access information for all IP SAN unit (4 of Fig. 2). Therefore, the distribution control management station (3 of Fig. 1) is able to accept the client hosts' (1 of Fig. 1) block data requests and to redirects these client hosts to each individual IP SAN Unit (4 of Fig. 1) for block data accessing.

f) When any IP SAN unit (n) shutdown, the IP SAN unit (n) (4 of Fig. 2) send "Unit (n) shutdown" to Distribution control Management station (3 of Fig. 2). This shutdown message could be an SNMP trap of link down, or a much simple UDP packet (Fig. 4) with message type of system down.
g) After received "unit (n) shutdown" packet from IP SAN unit (n) (4 of Fig. 2), the distribution control management modules on distribution control management station (3 of Fig. 2) updates and mark its IP SAN unit (n) status to down in a entry of the information list which corresponding to that IP SAN unit (n) (4 of Fig. 2).

When a IP SAN Unit (n) (4 of Fig. 2) system shutdown being automatically detected, the distribution control management station (3 of Fig. 2) also needs to update other information of the virtual SAN configuration, which related to a particular IP San Unit's (n) shutdown such as the total size of the virtual storage has as well as client hosts volume allocation information etc..

3: Distributed Virtual SAN volume Allocation and Access:

The Fig. 1 and Fig. 6 can explain how does multiple client hosts (1 of Fig. 1) access virtual SAN (Fig. 2) and how can they share an IP SAN unit (Fig. 1 & Fig. 5). The discussion of IP SAN data accessing will focus on how the storage requests being handled and how does the storage volume can be shared. Here, the term of storage volume is an abstract term rather than the actual term just for the convenience of the discussion. Further the very detailed steps of configuring actual volumes and partitions etc will be ignored in this invention. The following is an example of how does the volumes on a virtual SAN can be allocated and accessed by the client hosts.

Assuming an IP SAN unit 1 (1 of Fig. 1) has 200GB of storage space and it was configured with 4 volumes with 50GB, 50GB, 60GB, and 40GB respectively (Fig. 5). Also, assuming the IP SAN unit 2 (1 of Fig. 1) has 300GB capacity and it was configured with 3 volumes with 100GB each (Fig. 5). In addition, assuming that there were two client hosts made storage volume requests to distribution control management station, (3 of Fig. 1) the following actions will be taken:

- Client host 1(1 of Fig.1) requests two different size of storage volumes such as one with size of 50GB and another with size of 100GB. Distribution control management station (3 of Fig. 1) assign one of 50 GB volume on IP SAN unit 1 (4 of Fig. 1) and one of 100GB volume on IP SAN unit 2 (4 of Fig. 1) to client host 1(1 of Fig. 1).
- Client hosts 2 (1 of Fig. 1) requests one volume with size of 100GB. Distribution control management station (3 of Fig. 1) assigns another 100GB volume on IP SAN unit 2 (4 of Fig.1) to client host 2 (1 of Fig. 1).

The distribution control management station's (3 of Fig. 1) storage volumes assignment includes passes the designated IP address, volume number, size, start LBA and end LBA of IP SAN unit (i) (4 of Fig. 1) to the client host (1 of Fig. 1). Distribution control management station (3 of Fig. 1) also passes client host's (1 of fig. 1) IP address and the assigned volume number to the IP SAN unit (i) (4 of Fig. 1). Therefore, distribution control management station (3 of Fig. 1), IP SAN unit (4 of Fig. 1), and client hosts (1 of Fig. 1) are synchronized for the volumes assignment and client host mapping information. After obtained IP address and volume information of an IP SAN unit (4 of Fig. 1), the client hosts (1 of Fig. 1) can establish a direct data path to IP SAN unit (4 of Fig. 1) and directly access volumes on IP SAN unit (4 of Fig. 1) without further involvement of the distribution control management station (3 of Fig. 1). The results of above storage volume requests, allocation and accessing are shown in **Fig. 6**. It is clear that the IP SAN unit (2) has being shared by both client host (1) and client host (2).

The claims are:

- **A Method of Building, Automatic Configuring Distributed Virtual SAN in a Cross Network Domain Environment and Virtual SAN Storage Allocation and Accessing..**

1: An IP based Distributed Virtual SAN can be built up by using
   a) Multiple IP SAN units.